



Revista Micaela

ISSN: 2955-8646 (en línea) / 2709-8990 (Impresa)
Universidad Nacional Micaela Bastidas de Apurímac
Vice Rectorado de Investigación – Perú

Vol. 5 Num. 2 (2024) - Publicado: 22/03/24
<https://doi.org/10.57166/micaela.v5.n2.2024>
Páginas: 40 - 47

Recibido 20/08/2024; Aceptado 04/09/2024

<https://doi.org/10.57166/micaela.v5.n2.2024.156>

Autores:

1. **ORCID ID** <https://orcid.org/0009-0009-0696-7654>
Bryan Hurtado-Delgado, Egresado en Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac, Perú.
191215@unamba.edu.pe
2. **ORCID ID** <https://orcid.org/0009-0002-3242-8688>
Marycielo Xiomara Oscco-Guillen, Egresada en Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac, Perú.
191226@unamba.edu.pe
3. **ORCID ID** <https://orcid.org/0000-0002-2552-5669>
Mario Aquino-Cruz, Docente en la Universidad Nacional Micaela Bastidas de Apurímac Perú, MSc. en Informática, investigador en las áreas de informática educativa, IoT, inteligencia artificial y ciberseguridad.
maquino@unamba.edu.pe

Eficiencia del modelo de red neuronal convolucional DenseNet para la detección de somnolencia en conductores

Efficiency of the DenseNet convolutional neural network model for detecting drowsiness in drivers

Bryan Hurtado-Delgado¹, Marycielo Xiomara Oscco-Guillen² y Mario Aquino-Cruz³

Resumen. La somnolencia en conductores es una condición crítica que puede comprometer la seguridad vial, por lo que es esencial contar con métodos precisos para detectarla. Este estudio evaluó la arquitectura DenseNet121 para detectar somnolencia en conductores, centrándose en el estado de los ojos (abiertos o cerrados) y la manifestación de bostezos. El diseño del estudio fue no experimental, de nivel descriptivo y con enfoque cuantitativo. Se utilizó un conjunto de 2090 imágenes de conductores, obtenidas de Kaggle, que fueron preprocesadas con la biblioteca MediaPipe para facilitar la detección de rostros. El modelo DenseNet121 logró una precisión del 98.46% para el estado de bostezo y del 99.62% para el estado de los ojos. La matriz de confusión mostró clasificaciones perfectas en ambas categorías. El reporte de clasificación destacó un F1-Score, recall y precisión de 1.00, evidenciando su capacidad para clasificar correctamente todos los ejemplos. En las pruebas en tiempo real, el modelo mostró confianza del 87% al 97% en el bostezo y una consistencia del 99% en los ojos, aunque con algunas fallas. Estos hallazgos resaltaron la eficiencia del modelo en la detección de somnolencia, sugiriendo su potencial como herramienta valiosa para la seguridad vial.

Palabras Clave: DenseNet, Detección, Redes neuronales, Somnolencia.

Abstract. Drowsiness in drivers is a critical condition that can compromise road safety, so it is essential to have accurate methods to detect it. This study evaluated the DenseNet121 architecture to detect drowsiness in drivers, focusing on the state of the eyes (open or closed) and the manifestation of yawning. The study design was non-experimental, descriptive and quantitative in approach. A set of 2090 images of drivers, obtained from Kaggle, was used, which were preprocessed with the MediaPipe library to facilitate face detection. The DenseNet121 model achieved an accuracy of 98.46% for the yawning state and 99.62% for the eye state. The confusion matrix showed perfect classifications in both categories. The classification report highlighted an F1-Score, recall and precision of 1.00, evidencing its ability to correctly classify all examples. In real-time testing, the model showed 87% to 97% confidence in yawning and 99% consistency in eyes, albeit with some flaws. These findings highlighted the model's efficiency in detecting drowsiness, suggesting its potential as a valuable tool for road safety.

Keywords: DenseNet, Detection, Drowsiness, Neural Networks.

1 Introducción

La somnolencia es un estado en el que las personas pueden quedarse dormidas involuntariamente, afectando su capacidad de realizar actividades cotidianas de manera segura. Esta condición puede ser causada por diversos factores, como la falta de sueño, estrés, depresión, uso de ciertos medicamentos o trastornos del sueño, entre ellos la apnea del sueño y la narcolepsia [1], este estado es un grave peligro para quienes manejan, ya que afecta su capacidad de reacción y aumenta la posibilidad de accidentes, especialmente en viajes largos o durante la noche. Además, pone en riesgo no solo la vida del conductor, sino también la de los pasajeros y otros conductores en la vía [2]. La Organización Mundial de la Salud (OMS) señala que cada año, los accidentes en las vías provocan alrededor de 1.19 millones de muertes, con una tasa de mortalidad en hombres que es tres veces superior a la de las mujeres [3]. Según la Dirección General de Tráfico (DGT) en España, la somnolencia contribuye, de manera directa o indirecta, a entre el 15% y el 30% de los accidentes automovilísticos [4]. En Brasil, la Dirección de Tránsito del Estado de São Paulo (Detran.SP) alertó que hasta el 20% de los accidentes de tráfico están directamente relacionados con la somnolencia, evidenciando la gravedad de este problema en las carreteras brasileñas [14]. En cuanto a la situación en Perú, el Instituto Nacional de Estadística e Informática (INEI) reveló que en 2016 hubo un total de 116,659 accidentes de tránsito, de los cuales el 0.97% (1,131 accidentes) fueron causados por la fatiga del conductor. Lima fue la región más afectada con 813 accidentes, seguida por Puno con 38 y Arequipa con 34 [5]. En los últimos años, las redes neuronales convolucionales (CNN) han revolucionado el reconocimiento visual. Estas redes, organizadas en capas jerárquicas que aplican convoluciones, han mostrado un rendimiento superior a los modelos tradicionales como la bolsa de palabras visuales (BoVW) [6]. Una de las arquitecturas más innovadoras en este contexto es DenseNet, que se caracteriza por un patrón de conectividad único: cada capa está conectada a todas las demás dentro de un bloque denso. Esto permite que cada capa acceda a los mapas de características de las capas anteriores, fomentando la reutilización intensiva de características y resultando en un modelo más compacto y menos propenso al sobreajuste. Además, cada capa recibe supervisión directa de la función de pérdida a través de caminos de atajo, lo que proporciona una supervisión profunda implícita. Estas características hacen que DenseNet sea una opción ideal para problemas de predicción por píxel, y se ha demostrado que extiende su uso a la predicción del flujo óptico con un rendimiento comparable a los enfoques supervisados [7]. En este contexto, investigaciones recientes han profundizado en la aplicación de la inteligencia artificial para mitigar accidentes relacionados con la somnolencia en los conductores. Un ejemplo, es el trabajo de Ahmed et al. [8], quienes implementaron un esquema de ensemble learning utilizando dos modelos InceptionV3. Su enfoque se centró en desarrollar un sistema que identificara características faciales, específicamente ojos y boca, mediante la detección previa de estas áreas utilizando Multi-task Convolutional Neural Network (MT-CNN). Utilizaron imágenes de National Tsing Hua University (NTHU) dataset, combinaron las salidas de las dos InceptionV3 a través de una red neuronal feed-forward, alcanzando así una precisión del 97.23%. Esto evidenció que la arquitectura de ensemble fue más efectiva que modelos individuales, mostrando un rendimiento notable en la detección de somnolencia. Por otro lado, Namburi et al. [9] crearon un modelo que combinaba CNN y Long Short-Term Memory (LSTM) con el propósito de identificar con precisión la somnolencia y distracciones en conductores a partir del análisis de secuencias de video. Para ello, utilizaron un conjunto de 65 videos segmentados en ventanas de 16 fotogramas, donde la extracción de características faciales se realizó mediante FaceMesh, calculando el Eye Aspect Ratio (EAR) y el Mouth Aspect Ratio (MAR). Al implementar un modelo de Redes Neuronales Convolucionales y de Memoria a Largo Plazo (CNN-LSTM) que analizaba datos secuenciales, lograron una precisión del 93.60%, junto con métricas adicionales de 93.61% de precisión, 91.92% de recall y 92.68% de F1-score. Este estudio demostró que su enfoque era altamente efectivo para mejorar la seguridad de los conductores.

En otro estudio, Lim et al. [10] emplearon tres modelos CNN individuales, cada uno con diferentes cantidades de capas convolucionales y funciones de activación ReLU, para detectar somnolencia mediante el análisis de características faciales. Su objetivo fue crear un sistema de alerta en tiempo real que previniera accidentes causados por la somnolencia. Utilizando el MRL Eye Dataset (76,598 imágenes) y el Yawn Dataset (5,052 imágenes), aplicaron un enfoque de ensemble learning que combinó las salidas de las tres CNNs, logrando una precisión del 97.4% para los ojos y del 96.5% para la boca. Esto indicó que su enfoque de ensemble proporcionó una alta precisión en la detección de somnolencia. Mohanty et al. [11] desarrollaron un sistema de detección de somnolencia en tiempo real usando el modelo Dlib, con un detector facial Histograma de Gradientes Orientados (HOG) y 68 puntos faciales. El objetivo de su investigación fue prevenir accidentes monitoreando el cierre de ojos y los bostezos. Usaron los conjuntos de datos MRL Eye (84,898 imágenes) y YawDD (29 videos), calculando el EAR para detectar parpadeos y el MAR para bostezos. El sistema logró una precisión del 93.25% en detección de ojos y 96.71% en bostezos con el dataset, y 82.02% y 85.44% en pruebas reales. Concluyeron que funcionaba bien con buena iluminación, pero requería mejoras en condiciones variables.

Finalmente, Thampi et al. [12] desarrollaron un modelo CNN diseñado para detectar simultáneamente el movimiento de los ojos y el bostezo con el fin de prevenir accidentes por somnolencia. Su arquitectura incluyó tres capas de convolución 2D, max pooling y una capa de dropout. Utilizaron el MRL Eye Dataset (48,000 imágenes) y el YawnDD (2,900 imágenes).

nes), y entrenaron su modelo para clasificar estados de los ojos y bostezos. Su metodología de detección logró una precisión del 84.53% en el MRL Eye Dataset y del 96.42% en el YawDD, lo cual demostró que la combinación de técnicas de visión por computadora y CNN era eficaz en la detección de somnolencia.

El objetivo de este proyecto de investigación fue desarrollar y evaluar un modelo de red neuronal convolucional (CNN), específicamente DenseNet121 [12], para la detección de somnolencia en conductores. Se buscó identificar y clasificar señales de somnolencia, tales como el estado de los ojos (abiertos o cerrados) y la manifestación de bostezos, utilizando imágenes de conductores en diferentes estados de acuerdo con el dataset disponible en Kaggle [15].

2 Método

2.1 Tipo y nivel

La investigación se enmarcó en el tipo aplicada, centrada en la implementación de la arquitectura DenseNet121 para detectar somnolencia en conductores. Fue de nivel descriptivo, ya que evaluó el rendimiento de la arquitectura en términos de precisión y eficacia.

2.2 Diseño del estudio

El estudio empleó un diseño no experimental, caracterizado por la ausencia de manipulación de la variable independiente, enfocándose en el rendimiento de la arquitectura DenseNet121 en condiciones controladas. Se aplicó un diseño de corte transversal con un enfoque cuantitativo, recopilando datos en un único momento, empleando métricas numéricas para evaluar la efectividad de la arquitectura.

2.3 Muestra

Selección de la muestra. Se empleó un conjunto de datos disponible públicamente en la plataforma Kaggle, que contenía imágenes de ojos abiertos (Open), ojos cerrados (Closed), así como de conductores bostezando (yawn) y no bostezando (no_yawn). Los criterios de inclusión establecieron que las imágenes debían tener una calidad aceptable para el procesamiento mediante redes neuronales.

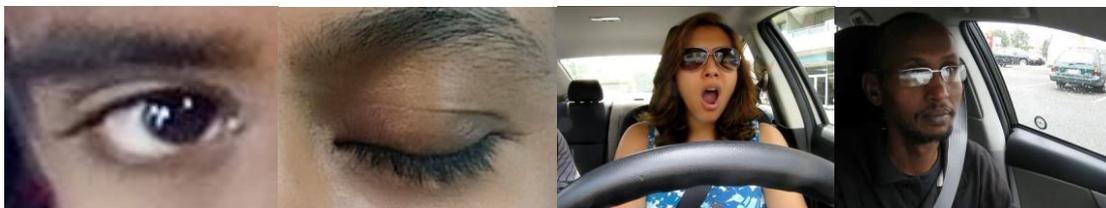


Fig. 1. Imágenes del conjunto de datos: categorías en Open, Closed, yawn y no_yawn.

Tamaño de la muestra. El conjunto de datos constó de un total de 2090 imágenes a color (RGB), distribuidas de la siguiente manera: 726 imágenes de ojos cerrados, 726 de ojos abiertos, 725 de conductores sin bostezar y 723 de conductores bostezando.

2.4 Diagrama de flujo

En la figura 2, se muestra un diagrama de flujo que detalla el procedimiento utilizado en el proyecto.

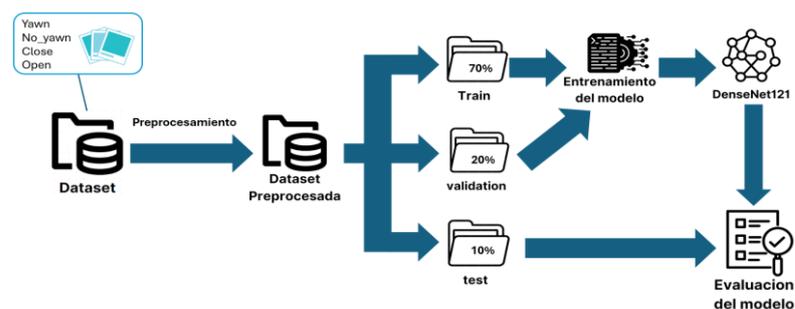


Fig. 2. Diagrama de flujo del procedimiento usado en la investigación. Elaboración propia.

2.5 Arquitectura CNN

Se presenta un esquema visual de la arquitectura empleada, en la figura 3, destacando las diferentes capas y sus interconexiones, lo que ilustra la complejidad y funcionalidad de esta red neuronal.

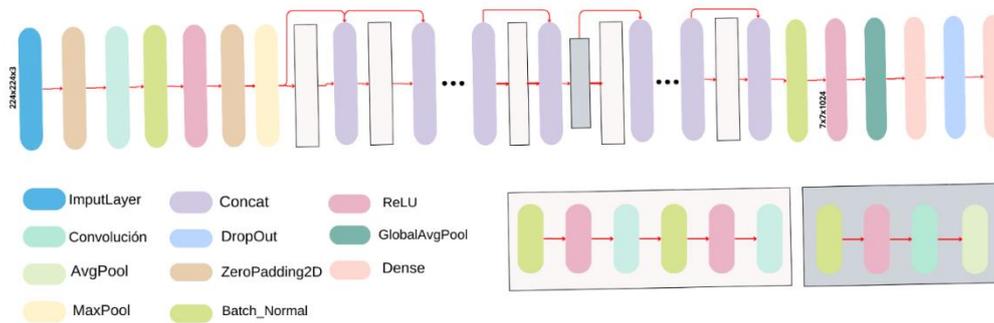


Fig. 3. Imagen de la arquitectura DenseNet121. Elaboración propia.

2.6 Procedimiento

- El dataset se dividió en dos partes: Estado de los ojos (closed y open) y estado de bostezo (yawn y no_yawn).
- Se llevó a cabo un proceso de preparación de datos que implicó ajustar los valores de los píxeles para que se encontraran en un rango de 0 a 1. Las imágenes se redimensionaron a 224x224 píxeles y se agruparon en conjuntos de 64. Además, se clasificaron de manera binaria, dividiendo los estados de los ojos en “open” y “closed”, y los estados de bostezo en “yawn” y “no_yawn”.
- Se realizó un preprocesamiento complementario específico para detectar el estado de bostezo, que consistió en aplicar una malla (landmark) sobre los rostros de los conductores. Este enfoque se implementó con el objetivo de mejorar la capacidad del modelo para identificar rostros de tonalidad más oscura y aquellos que usaban lentes, ya sean oscuros o transparentes. Para llevar a cabo este proceso, se empleó la biblioteca MediaPipe y su función FaceMesh, lo que permitió recortar la imagen alrededor del landmark y centrarse en la cara del conductor.

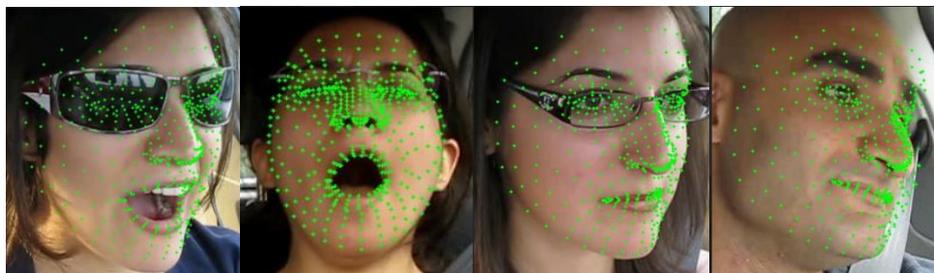


Fig. 4. Imágenes del estado de bostezo ya preprocesadas

- Para el estado de los ojos, se dividió el conjunto de datos en un 70% para entrenamiento (1,060 imágenes), un 20% para validación (264 imágenes) y un 10% para pruebas (148 imágenes). De manera similar, para el estado de bostezo, el conjunto de datos se separó en un 70% para entrenamiento (1,042 imágenes), un 20% para validación (260 imágenes) y un 10% para pruebas (146 imágenes).
- Se llevó a cabo el entrenamiento de la red neuronal convolucional DenseNet121, partiendo de un modelo preentrenado con los pesos de ImageNet. El proceso se desarrolló en Google Colab bajo una cuenta estándar, lo que, aunque con ciertas limitaciones de recursos, permitió finalizar el entrenamiento de manera efectiva. Se utilizó el optimizador Adam con una tasa de aprendizaje de 0.001, y se aplicó la función de pérdida binary_crossentropy por su adecuación a la clasificación binaria. Además, se implementó la técnica de regularización mediante la penalización L2 (kernel_regularizer=l2(0.01)) en las capas fully connected y se añadió una capa dropout del 50% para prevenir el sobreajuste (overfitting). Durante cada época, el rendimiento del modelo fue evaluado utilizando la métrica de precisión (accuracy).
- Para optimizar el proceso de entrenamiento, se implementó la técnica de Early Stopping, para detener el entrenamiento de manera anticipada si la pérdida en el conjunto de validación no mejoraba después de cinco épocas consecutivas y en caso se detenía, se restauraban los pesos que ofrecieron el mejor desempeño.
- El tiempo de entrenamiento fue de 9 horas aproximadamente para el entrenamiento del estado de bostezo y casi 10 horas para el entrenamiento del estado de los ojos.

h) Posteriormente, se llevó a cabo la fase de pruebas utilizando las imágenes destinadas para el testeo. El modelo entrenado fue guardado en formato .h5 para facilitar su evaluación y uso posterior. Finalmente, se desarrolló un prototipo sencillo en Google Colab que activa la cámara y procesa el vídeo en tiempo real. Cada fotograma capturado se preprocesa y se envía al modelo entrenado, permitiendo mostrar de manera inmediata el estado del conductor sobre la imagen de la cámara.

2.7 Instrumentos y materiales

Con el fin de realizar esta investigación, se optó por utilizar Google Colab como plataforma de desarrollo, haciendo uso de Python como lenguaje de programación. Se integraron diversas bibliotecas, incluyendo TensorFlow, Keras, Matplotlib, MediaPipe y OpenCV, para implementar la arquitectura DenseNet121. El conjunto de datos, disponible en Kaggle, contenía imágenes organizadas que facilitaban la detección de somnolencia en conductores. Asimismo, se recurrió a Jira para la gestión del artículo, implementando la metodología Scrum para estructurar y supervisar el avance del estudio de manera eficiente.

2.8 Análisis de datos

Curvas de precisión y pérdida. Se generaron gráficos que mostraron la evolución de la precisión y la pérdida en los conjuntos de entrenamiento y validación, permitiendo identificar el comportamiento de DenseNet121 y detectar posibles problemas de sobreajuste.

Precisión en validación y entrenamiento. Se calculó la precisión en porcentaje para ambos conjuntos, ofreciendo una medida cuantitativa de la efectividad del modelo en la predicción de estados de bostezo y ojos abiertos, lo que refleja su capacidad de aprendizaje y generalización.

Matriz de confusión. Se utilizaron matrices para analizar aciertos y errores en la clasificación, identificando verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN), lo que fue crucial para evaluar el rendimiento del modelo y mejorar su discriminación entre clases.

Reporte de clasificación. Se elaboró un informe con métricas de precisión, recall y F1-score, evaluando el desempeño del modelo. Se calcularon promedios y la exactitud general, permitiendo analizar su capacidad para diferenciar entre clases y gestionar casos de verdaderos y falsos positivos, lo que respalda la efectividad de DenseNet121 en la detección de somnolencia.

3 Resultados

3.1 Curvas de entrenamiento y validación

La figura 5 muestra las curvas de precisión y pérdida para la arquitectura DenseNet121, correspondientes a las fases de entrenamiento y validación en los estados de bostezo ('yawn' y 'no_yawn') y los estados de los ojos ('closed' y 'open'). Las curvas mostraron la evolución de la precisión y la pérdida a lo largo de 32 épocas para el estado de bostezo y 33 épocas para el estado de los ojos.

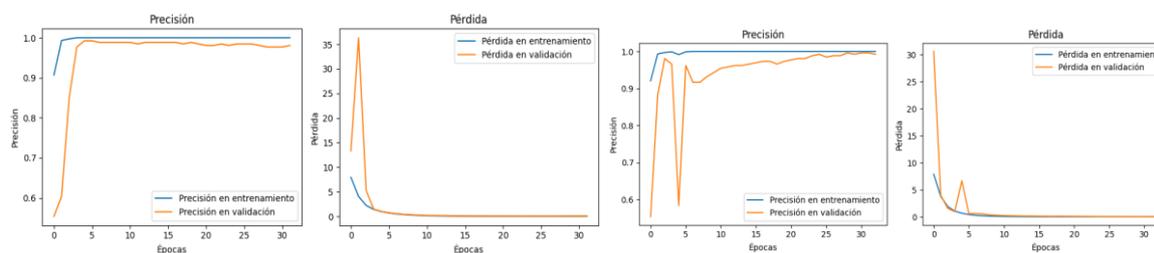


Fig. 5. Curvas de precisión y pérdida de DenseNet121 tras 32 épocas para el estado bostezo y 33 épocas para el estado de los ojos.

3.2 Precisión en validación y entrenamiento

La tabla 1 presenta la precisión del modelo evaluado para los estados de bostezo y ojos, en las etapas de entrenamiento y validación.

Tabla 1. Precisión en Entrenamiento y Validación para el estado de Bostezo y Ojos.

Arquitectura	Estado	Entrenamiento	Validación
DenseNet121	Bostezo	100.00%	98.46%
	Ojos	100.00%	99.62%

3.3 Matriz de confusión

Las matrices de confusión para la arquitectura DenseNet121, el cual se llevó a cabo con los datos de prueba (test), se muestran en la figura 7. En el estado de ojos, la matriz relevó 74 TP para ‘Closed’ y 74 TN para ‘Open’, sin FP ni FN. Para el estado de bostezo, se obtuvieron 73 TP para ‘yawn’ y 73 TN para ‘no_yawn’. también si FP ni FN.

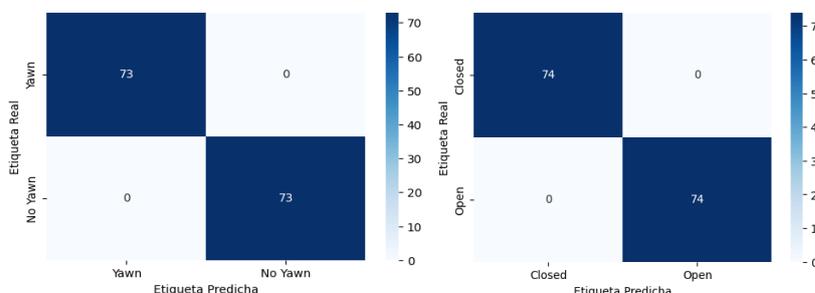


Fig. 6. Matrices de confusión para la arquitectura DenseNet121 para el estado de bostezo y ojos.

3.4 Reporte de clasificación

Las tablas 2 y 3, que también se llevó con los datos de prueba (test), presentaron las métricas de precisión, recall, F1-score y soporte para la arquitectura DenseNet121 en los estados de bostezo (‘yawn’ y ‘no_yawn’) y los estados de ojos (‘closed’ y ‘open’). Los valores de precisión, recall y F1-score fueron de 1.00 para todas las clases en ambos estados, con un soporte total de 146 para los estados de bostezo y 148 para los estados de ojos.

Tabla 2. Métricas y promedios para las clases de bostezo

Clase	Precisión	Recall	F1-Score	Support
yawn	1.00	1.00	1.00	73
no_yawn	1.00	1.00	1.00	73
Accuracy			1.00	146
Macro avg	1.00	1.00	1.00	146
Weighted avg	1.00	1.00	1.00	146

Tabla 3. Métricas y promedios para las clases de ojos

Clase	Precisión	Recall	F1-Score	Support
Closed	1.00	1.00	1.00	74
Open	1.00	1.00	1.00	74
Accuracy			1.00	148
Macro avg	1.00	1.00	1.00	148
Weighted avg	1.00	1.00	1.00	148

3.5 Pruebas en tiempo real del prototipo

El prototipo presentó una latencia variable en las predicciones, alcanzando en el peor de los casos hasta 5 segundos y en el mejor de los casos 0.5 segundos. En la mayoría de las pruebas, el promedio de latencia fue inferior a 2.75 segundos. Esta latencia se genera debido al procesamiento en tiempo real de cada fotograma capturado por la cámara, junto con la ejecución del modelo entrenado en Colab, lo que incluye tanto la etapa de preprocesamiento de las imágenes como la evaluación de la arquitectura. Además, factores como la velocidad de conexión a internet y la carga computacional de cada arquitectura influyen en estos tiempos.

Cabe recalcar que los datos fueron obtenidos mediante la observación constante del prototipo y cada predicción realizada por el algoritmo incluye un valor de confianza, el cual representa qué tan seguro está la arquitectura respecto a su predicción. Este valor se expresa como un porcentaje, donde una mayor confianza (cercano al 100%) significa que el modelo

está convencido de su clasificación, y un valor menor (cercano al 50% o menos) indica dudas en la predicción. Es importante señalar que, en situaciones donde una arquitectura arroje resultados incorrectos, pero presente un nivel de confianza cercano al 100%, carecería de transparencia.

Clasificación del estado de bostezo. El modelo entrenado presentó niveles de confianza sólidos, alcanzando entre el 87% y el 97% en sus predicciones, lo que indica una seguridad considerable en su capacidad para identificar el estado de bostezo en los conductores. Además, demostró transparencia en sus predicciones

Clasificación del estado de los ojos. El modelo entrenado logró una confianza constante del 99% con predicciones correctas, fallando algunas veces, pero mostrando una correcta transparencia en sus resultados.

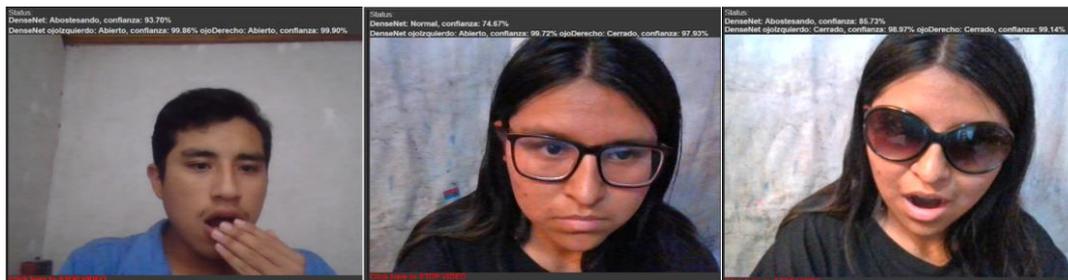


Fig. 7. Imágenes del prototipo ejecutado en tiempo real

4 Discusiones

En el estudio de Ahmed et al. [8], se utilizó un enfoque de ensemble learning con dos modelos InceptionV3, entrenados para extraer características faciales de los ojos y la boca, alcanzando una precisión del 97.03% en la detección de somnolencia. En comparación, DenseNet121, empleado en este estudio, obtuvo una precisión superior del 98.08% en la detección de bostezos y un 100% en la detección de ojos abiertos, demostrando su eficacia sin la complejidad de combinar múltiples modelos.

Por otro lado, Namburi et al. [9], aplicando un modelo CNN-LSTM secuencial, lograron una precisión de 93.60% en la detección de somnolencia. No obstante, su enfoque requería secuencias de video y la extracción de características como el EAR y el MAR. Por otro lado, DenseNet121 trabajó con imágenes estáticas y logró mayores precisiones, simplificando el proceso de detección con un 98.08% en bostezos y 100% en los ojos.

De manera similar, el estudio de Lim et al. [10] empleó tres modelos especializados: uno para detectar ojos, otro más profundo también para ojos, y uno para detectar bostezos. Lograron un 97.4% de precisión en ojos y un 96.5% en bostezos. Sin embargo, los resultados de DenseNet121 que se realizó en el presente estudio fueron superiores, alcanzando un 100% en la detección de ojos abiertos y un 98.08% en bostezos, utilizando solo un modelo.

Asimismo, en el trabajo de Mohanty et al. [11], se desarrolló un sistema basado en Dlib y HOG, logrando 96.71% en la detección de bostezos y 93.25% en ojos cerrados en condiciones controladas. Sin embargo, su precisión disminuyó a 85.44% y 82.02% en entornos reales con variaciones de iluminación. En cambio, DenseNet121 mantuvo una alta precisión, con un 98.08% en bostezos y un 100% en ojos.

Finalmente, Thampi et al. [12] usaron una arquitectura CNN de tres capas de convolución 2D y alcanzaron 84.53% de precisión en el MRL Eye Dataset y 96.42% en el YawDD. Aunque obtuvieron buenos resultados, la arquitectura DenseNet121 en este estudio mostró una precisión superior, especialmente en ojos con una precisión de 100% y bostezos con 98.08% de precisión, evidenciando su capacidad para extraer características faciales más detalladas y relevantes.

5 Conclusiones y recomendaciones

En este estudio, se evaluó y se utilizó un modelo de red neuronal convolucional DenseNet121 para la detección de somnolencia en conductores, enfocándose en dos aspectos fundamentales: el estado de los ojos y el estado de bostezo. Para la clasificación del estado de bostezo, el modelo entrenado demostró un rendimiento sobresaliente con una precisión en validación de 98.46%. Para la clasificación del estado de los ojos también se destacó obteniendo una precisión en validación del 99.62%.

Es relevante señalar que el entrenamiento del modelo se realizó utilizando solo la CPU, ya que el uso de GPU cortaba drásticamente el tiempo de uso, para un usuario sin suscripción. La decisión de usar la CPU permitió una mayor continuidad en el entrenamiento, sin interrupciones. Fuera de eso, es importante detallar que la memoria RAM alcanzaba su límite máximo durante el entrenamiento de DenseNet121 para el estado de bostezo, lo que provocaba que el entrenamiento se detuviera inesperadamente. Para ello se implementó un mecanismo de checkpoint únicamente para este caso, el cual guardaba los pesos del modelo al alcanzar la mejor pérdida de validación, permitiendo retomar el entrenamiento desde ese punto óptimo sin perder progreso.

En conclusión, el estudio resalta la eficacia del modelo DenseNet121 en la detección de somnolencia en conductores, mostrando un excelente rendimiento en la clasificación del estado de bostezo. Sin embargo, en el estado de los ojos, se observaron algunas fallas en el prototipo durante las pruebas en tiempo real. A pesar de esto, los resultados generales son confiables y consistentes, lo que evidencia el potencial del modelo como una herramienta valiosa para mejorar la seguridad vial al monitorear el estado de alerta de los conductores.

6 Biografías

- Bryan Hurtado Delgado, Egresado en Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac.
- Marycielo Xiomara Oscco Guillen, Egresada en Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac.
- Mario Aquino Cruz, Docente en la Universidad Nacional Micaela Bastidas de Apurímac Perú, MSc. en Informática, investigador en las áreas de informática educativa, IoT, inteligencia artificial y ciberseguridad.

7 Referencias

- [1] National Library of Medicine, "Drowsiness," MedlinePlus Medical Encyclopedia. [Accedido: 20 de septiembre de 2024]. [En línea]. Disponible: <https://medlineplus.gov/ency/article/003208.htm>
- [2] K. Peña Prado, "Somnolencia en conductores de transporte público regular de pasajeros de Lima Metropolitana – Perú," Tesis de Maestría, Universidad Peruana Cayetano Heredia, Lima, Perú, 2017.
- [3] World Health Organization, "Road traffic injuries", 2023. [Accedido: 22 de septiembre de 2024]. [En línea]. Disponible: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>
- [4] Dirección General de Tráfico: "Conducir con sueño o cansancio." *DGT*, 2023. [Accedido: 23 de septiembre de 2024]. [En línea]. Disponible: <https://www.dgt.es/muevete-con-seguridad/evita-conductas-de-riesgo/Conducir-con-sueno-o-cansancio>
- [5] Instituto Nacional de Estadística e Informática: "Análisis de los Accidentes de Tránsitos Ocurredos en el Año 2016.", 2017. [Accedido: 24-sep-2024]. [En línea]. Disponible: <https://www.inei.gov.pe/est/lib1528/cap03>
- [6] Xie and A. Yuille, "Genetic CNN" en *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Baltimore, MD, USA, 2017, pp. 1379-1386.
- [7] Y. Zhu y S. Newsam, "DenseNet for dense flow," en *IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 999-1003, doi: 10.1109/ICIP.2017.8296389.
- [8] Ahmed, T., Jyoti, O., Mou, T.H.: Drivers' drowsiness detection system enhancement using deep learning: CNN-based approach. In: *2023 26th International Conference on Computer and Information Technology (ICCIT)*, pp. 1–6. IEEE, Cox's Bazar, Bangladesh (2023). doi: 10.1109/ICCIT60459.2023.10441229.
- [9] Namburi, A., Sitpasert, P., Duang-onnam, W.: A CNN-LSTM approach for accurate drowsiness and distraction detection in drivers. *ICIC Express Letters* 18, 907–917 (2024). doi: 10.24507/icicel.18.09.907.
- [10] Lim, B.-E., Ng, K.-W., Ng, S.: Drowsiness detection system through eye and mouth analysis. *JOIV: International Journal on Informatics Visualization* 7, 2489 (2023). doi: 10.62527/joiv.7.4.2288.
- [11] Mohanty, S., Hegde, S.V., Prasad, S., Manikandan, J.: Design of real-time drowsiness detection system using Dlib. In: *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, pp. 1–5 (2019). doi: 10.1109/wiecon-ece48653.2019.9019910.
- [12] Thampi, L., Kashyap, K., T, N., Reddy, A., Khan, I., Aswathy, M., Kumar, A., Kumar, S.: Smart Driver Assistance: Real-Time Drowsiness Detection Using CNN and Computer Vision. *Preprint, Research Square* (2024). <https://doi.org/10.21203/rs.3.rs-4962655/v1>.
- [13] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708 (2017).
- [14] São Paulo Governo do Estado São Paulo São Todos, "Dirigir com sono causa 20% dos acidentes de trânsito | Governo do Estado de São Paulo", *Governo do Estado de São Paulo*, Apr. 15, 2017. [Accedido: 25 de septiembre de 2024]. [En línea]. Disponible: <https://www.saopaulo.sp.gov.br/spnoticias/dirigir-com-sono-causa-20-acidentes/>.
- [15] Dheeraj Perumandla, "Drowsiness_dataset," Kaggle, 2020. [Accedido: 20 de septiembre de 2024]. [En línea]. Disponible: <https://www.kaggle.com/datasets/dheerajperumandla/drowsiness-dataset>